

Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags

I. Coulibaly^{*,†}, K. Gharbi[‡], R. G. Danzmann[‡], J. Yao^{*} and C. E. Rexroad III[†]

^{*}Animal and Veterinary Sciences Division, West Virginia University, PO Box 6108, Morgantown, WV 26506, USA. [†]National Center for Cool and Cold Water Aquaculture, USDA-ARS, Leetown, WV 25430, USA. [‡]Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1 Canada

Summary

The construction of high-density linkage maps for use in identifying loci underlying important traits requires the development of large numbers of polymorphic genetic markers spanning the entire genome at regularly spaced intervals. As part of our efforts to develop markers for rainbow trout (*Oncorhynchus mykiss*), we performed a comparison of allelic variation between microsatellite markers developed from expressed sequence tag (EST) data and anonymous markers identified from repeat-enriched libraries constructed from genomic DNA. A subset of 70 markers (37 from EST databases and 33 from repeat enriched libraries) was characterized with respect to polymorphism information content (PIC), number of alleles, repeat number, locus duplication within the genome and ability to amplify in other salmonid species. Higher PIC was detected in dinucleotide microsatellites derived from ESTs than anonymous markers (72.7% vs. 54.0%). In contrast, dinucleotide repeat numbers were higher for anonymous microsatellites than for EST derived microsatellites (27.4 vs. 18.1). A higher rate of cross-species amplification was observed for EST microsatellites. Approximately half of each marker type was duplicated within the genome. Unlike single-copy markers, amplification of duplicated microsatellites in other salmonids was not correlated to phylogenetic distance. Genomic microsatellites proved more useful than EST derived microsatellites in discriminating among the salmonids. In total, 428 microsatellite markers were developed in this study for mapping and population genetic studies in rainbow trout.

Keywords diversity, expressed sequence tag, microsatellites, polymorphism, salmonid, trout.

Introduction

The rainbow trout (*Oncorhynchus mykiss*) comprises both anadromous and non-anadromous fishes belonging to the Salmonidae family originating from the Pacific coast of North America and Russia. Introductions have extended its distribution to many countries around the world in regions with cool water temperatures (MacCrimmon 1971) making it one of the most widely cultivated food and sport fishes. The development of molecular markers has enabled population genetic studies determining population structures and

interactions (Nielsen *et al.* 1999; Heath *et al.* 2002; Ardren & Kapuscinski 2003).

Most microsatellites markers are neutral or Type II markers (O'Brien *et al.* 1993), although tandem nucleotide repeats are present in coding regions of the genome. Because of their location in well-conserved regions of the genome, cDNA derived microsatellites are expected to be conserved across closely related species and represent a potential source of Type I markers (O'Brien *et al.* 1993). Rainbow trout cDNAs and expressed sequence tags (ESTs) available in public databases offer an *in silico* approach to marker development at virtually no cost. Searches of EST databases for microsatellite containing sequences have been useful for a number of species including humans (Haddad *et al.* 1997), catfish (Serapion *et al.* 2004), rice (Cho *et al.* 2000) and barley (Thiel *et al.* 2003). Marker development in salmonids is complicated by the evolutionarily recent

Address for correspondence

Caird Rexroad, National Center for Cool and Cold Water Aquaculture, 11861 Leetown Road, Kearneysville, West Virginia 25430, USA.
E-mail: crexroad@nccwa.ars.usda.gov

Accepted for publication 31 March 2005

genome duplication event which often results in multiple copies of loci in the haploid genome (Venkatesh 2003). Characterization of loci including copy number is important when conducting analyses of genetic variability in genomic regions under control of different evolutionary constraints.

Two linkage maps have been published for rainbow trout using AFLP and microsatellite markers with an average marker spacing of 10 cM (Sakamoto *et al.* 2000; Nichols *et al.* 2003). For effective use of genomic tools in rainbow trout breeding, it is critical to develop novel markers useful for the construction of high-density linkage maps. In the study presented here, we report on the development of anonymous rainbow trout microsatellites from repeat-enriched libraries and from ESTs retrieved from a public database and their cross-species amplification potential in nine other salmonids. We also investigated sequence variation in terms of polymorphism, number of alleles and allele repeat count to address the relative informativeness of both types of markers. Marker potential for genetic diversity analyses among the salmonids was also assessed.

Materials and methods

Identification of microsatellites

Three genomic libraries enriched for ATG, CA and GA microsatellite repeats, respectively, were acquired from Genetic Identification Services (Chatsworth, CA, USA). Fragments 350–700 bp in size were cloned into the pUC19 vector and transformed into DH5 α cells to generate the library. The libraries were arrayed into 96-well plates for bidirectional sequencing on an ABI3700 automated DNA sequencer using ABI Prism Big Dye Terminator cycle sequencing kit (ABI, Foster City, CA, USA) and M13 forward and reverse primers. Redundant and overlapping sequences were grouped using Vector NTI Suite 7.0 software (Informax, Bethesda, MD, USA). Unique sequences were compared by BLAST (Altschul *et al.* 1990) against Genbank non-redundant STS and EST databases prior to primer design in order to analyse for redundancy with previously published markers and to identify matches from coding sequences.

The rainbow trout gene index (RTGI, http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=r_trout) database of The Institute for Genomic Research was searched for microsatellite-repeat-containing sequences using Tandem Repeat Finder v 3.21 software (Benson 1999). A total 21 303 tentative consensus sequences were analysed. Only sequences containing di, tri, tetra and penta-nucleotides with at least 12, 8, 6 and 5 unit repeats, respectively, were selected for marker development.

Polymerase chain reaction (PCR) primer pairs were designed to amplify products from 100 to 300 bp using

Oligo 6.0 software (Rychlik & Rhoads 1989). Forward primers of each pair were commercially labelled either by addition of M13 tail (Boutin-Ganache *et al.* 2001) or by addition of FAM, HEX or NED fluorescent dyes at the 5' end of the oligonucleotides. PCR optimization was performed by varying annealing temperatures and MgCl₂ concentrations. PCR amplifications were conducted in a 12 μ l reaction volume containing 50 ng DNA and 12.5 ng DNA, 1.5–2.5 mM MgCl₂, 1.0 μ M of each primer, 200 μ M of dNTPs, 1X manufacturer's reaction buffer and 0.5 unit Taq DNA polymerase. The PCR programme consisted of an initial denaturation at 95 °C for 15 min followed by 30 cycles of 95 °C for 1 min, annealing temperature for 45 s, 72 °C extension for 45 s and a final extension at 72 °C for 10 min. PCR products were visualized on agarose gels after staining with ethidium bromide.

Microsatellite characterization

Microsatellite markers were amplified from a DNA panel representing 48 individuals. Twenty-four rainbow trout from laboratory broodstock (Silverstein *et al.* 2004) were used for polymorphism analyses along with five fish from doubled haploid lines (Arlee, Clearwater, Hot Creek, OSU, Swanson, Young *et al.* 1996) to assess locus duplication. Cross-species amplification was tested using 19 other salmonids: *Oncorhynchus clarki* Coastal cutthroat trout (3), *Oncorhynchus kisutch* Coho (2), *Oncorhynchus nerka* Sockeye (2), *Oncorhynchus tshawytscha* Chinook (2), *Salmo salar* Atlantic salmon (2), *Salmo trutta* Brown trout (2), *Salvelinus alpinus* Arctic char (2), *Salvelinus fontinalis* Brook trout (2) and *Thymallus arcticus* Arctic grayling (2). DNA samples were isolated from fin clips using the phenol–chloroform method (Sambrook & Russell 2001) and PCR conducted as described above. Amplicons were visualized on 3% agarose gels. Markers were grouped in combinations of two or three markers based on differences in dye and amplicon size. Three microlitres of each PCR product was added to 20 μ l water. One microlitre of the diluted sample was added to 12.5 μ l of loading mixture made up with 12 μ l HiDi formamide and 0.5 Genscan 400 ROX internal size standard. Samples were denatured at 95 °C for 5 min and kept on ice until loading on an automated DNA sequencer ABI 3730 DNA Analyzer (ABI). Output files were analysed using GeneMapper (ABI).

Data analyses

For each marker allelic variation was estimated by the polymorphism information content (PIC) value first described by Botstein *et al.* (1980) and modified by Anderson *et al.* (1993). Polymorphism information content was calculated as follows:

$$\text{PIC} = 1 - \sum_{i=1}^k p_i^2$$

where p_i is the frequency of the i th allele for a given microsatellite and k is the total number of alleles detected for that microsatellite. We estimated heterozygosity as the number of heterozygous genotypes divided by the total number of genotypes scored at a given locus, excluding the clonal lines.

The repeat count of markers individual alleles was estimated using the formula:

$$\frac{(As - Rs)}{Ru} + Rc$$

where As is the estimated allele size, Rs the reference allele size, i.e. the size of the amplicon in the strain from which the sequence was obtained, Ru is the repeat unit length (2, 3, 4 or 5), Rc is the reference allele repeat count manually determined from its nucleotide sequence. This formula assumes that allele size variation is due solely to microsatellite repeat variation and no indel occurred in the flanking sequences. This assumption is supported by a small rate of microsatellite size variation reported to be due to mutation in the flanking sequences in *Drosophila* (Colson & Goldstein 1999) and *Limulus polyphenus* (Ortí *et al.* 1997). When negative values of repeats count were found as a result of high difference in size between alleles and reference allele the value was dropped from further analyses. Markers containing two alleles in at least one rainbow trout double haploid and at least three peaks in diploid fish were determined to be duplicated in the genome. Statistica software version 6 (StatSoft Inc., Tulsa, OK, USA) was used for statistical analyses (ANOVA).

For the purpose of assessing genetic similarity, PCR products derived from EST-microsatellites and anonymous microsatellites were scored in binary format, with the presence of an allele scored as 1 and its absence scored as 0. Pair-wise similarity coefficients based on Jaccard's algorithm were computed and the similarity matrix thus obtained was used in Principal Coordinates (PCO) analysis using the PAST software (Hammer *et al.* 2001).

Results

Sequences from 2246 clones from repeat-enriched libraries and 21 303 transcribed sequences were searched for microsatellite repeats leading to the development of 249 and 179 novel microsatellite markers, respectively. All 428 markers including sequence, primer sequences, and PCR conditions were submitted to the STS database in GenBank and assigned accession numbers BV211862 through BV212280 and BV212286 through BV212294.

A subset of 70 microsatellites markers, 37 from repeat-enriched libraries and 33 from ESTs, were randomly selected for further study. No significant BLAST matches

with EST and nr databases were identified for sequences from the repeat-enriched libraries.

Polymorphism

Half of the 70 microsatellite markers analysed were found to be duplicated within the genome. Twenty were derived from repeat-enriched libraries and 15 were derived from ESTs. These markers were discarded from the polymorphism survey. Assignment of alleles to loci was too difficult because of a small range of allele sizes and a high frequency of null alleles.

Thirty-five single-copy microsatellite markers, 17 from repeat-enriched libraries (12 with di-nucleotide repeats and five with tri-nucleotide repeats) and 18 from ESTs (16 with dinucleotide repeats, one with tri-nucleotide repeats and one with penta-nucleotide repeat) were screened for informativeness. All markers showed polymorphism within rainbow trout panel except OMM1748 (Table 1). The average heterozygosity value was $52.0 \pm 27.4\%$ in dinucleotide EST-microsatellites and $46.6 \pm 28.5\%$ in dinucleotide anonymous microsatellites, the difference not being significant ($F_{1,26} = 0.3$; $P = 0.621$). The average PIC value and the average number of alleles were significantly higher for EST-microsatellites than for repeat-enriched library derived microsatellites ($72.7 \pm 16.5\%$ vs. $54.0 \pm 24.0\%$, $F_{1,26} = 6.0$ $P = 0.022$) and (8.6 ± 4.0 vs. 5.3 ± 3.3 , $F_{1,26} = 5.1$ $P = 0.033$), respectively.

A highly significant difference was found for average repeat number between anonymous (31.6 ± 11.3) and EST dinucleotide-repeat (16.5 ± 8.0) microsatellites ($F_{1,25} = 16.3$; $P = 0.000$). A large majority of genomic microsatellites (82%) had an average repeat number higher than that of the reference allele selected for microsatellite identification. The reverse situation was observed for EST-microsatellites as 62% of markers showed an average repeat count lower than that of the reference allele. In Fig. 1, the average repeat number is plotted against the average number of alleles for each marker of both types. The EST-derived OMM5188 was excluded from this analysis because the average repeat number was three times higher than reference allele repeat number. Both types of markers expressed wide ranges of variation for the number of alleles. For transcribed microsatellites the dispersion for repeat count is confined to a small amplitude while a wide dispersion was observed for genomic microsatellites.

Translated TC sequences containing the 18 single-copy EST microsatellites were BLASTed against the NCBI protein database. Alignment scores higher than 200 (201–1902) were identified for 15 markers. Of these, 14 microsatellite repeats were located downstream of the translated region of the TC sequence and one microsatellite repeat was located upstream the translated region, putatively in the 3'UTR (untranslated regions) and the 5'UTR respectively.

Locus name	GenBank acc no.	Marker source	Core repeat	No. alleles	Size range (bp)	PIC (%)
OMM1511	BV212047	MREL	(TC) ₂₅ *	15	114–274	87
OMM1516	BV212051	MREL	(TC) ₁₅	7	166–191	80
OMM1542	BV212071	MREL	(TC) ₁₃ *	3	133–187	16
OMM1559	BV212287	MREL	(GA) ₁₃	4	190–204	55
OMM1561	BV212083	MREL	(TC) ₉ *	3	215–249	19
OMM1579	BV212097	MREL	(TC) ₁₇	6	101–158	71
OMM1590	BV212106	MREL	(CAT) ₆ *	8	104–207	83
OMM1693	BV212179	MREL	(CA) ₁₃	5	156–201	67
OMM1696	BV212181	MREL	(CA) ₈ *	2	168–170	28
OMM1701	BV212184	MREL	(CAT) ₈	7	201–225	51
OMM1727	BV212207	MREL	(CAT) ₈	3	244–254	49
OMM1739	BV212214	MREL	(ATG) ₈ *	6	174–201	69
OMM1741	BV212216	MREL	(GA) ₉ *	5	231–252	65
OMM1748	BV212223	MREL	(GGA) ₄ *	1	247	0
OMM1770	BV212239	MREL	(TG) ₁₈	3	296–322	37
OMM1783	BV212249	MREL	(GA) ₇ *	3	213–251	51
OMM1832	BV212279	MREL	(TC) ₁₄ *	7	225–383	7
OMM5129	BV211862	EST	(GT) ₂₁	12	97–143	83
OMM5132	BV211863	EST	(CA) ₁₇ *	12	96–130	89
OMM5133	BV211864	EST	(AT) ₁₉	9	125–175	78
OMM5136	BV211865	EST	(AT) ₁₇	7	122–153	45
OMM5137	BV211866	EST	(CA) ₁₁ *	3	176–180	60
OMM5141	BV211870	EST	(AG) ₁₀ *	3	225–228	32
OMM5166	BV211894	EST	(AT) ₁₆	4	133–170	63
OMM5177	BV211903	EST	(TGA) ₉	10	112–151	85
OMM5188	BV211912	EST	(AG) ₁₅ *	11	160–246	78
OMM5206	BV211928	EST	(GT) ₁₄ *	11	135–178	86
OMM5233	BV211954	EST	(GA) ₁₇	13	111–137	87
OMM5255	BV211970	EST	(AT) ₁₇	13	208–255	85
OMM5258	BV211973	EST	(AT) ₂₀	19	130–182	90
OMM5264	BV211979	EST	(GT) ₁₃ *	11	98–114	79
OMM5265	BV211980	EST	(CTCCT) ₄ *	7	143–153	57
OMM5267	BV211982	EST	(GT) ₁₅ *	8	144–171	62
OMM5268	BV211983	EST	(GT) ₁₅ *	10	123–160	75
OMM5295	BV212005	EST	(TA) ₂₁ *	8	125–236	71

Marker source is denoted as MREL for microsatellite repeat-enriched library and EST for expressed sequence tag.

*Imperfect repeat.

Cross-species amplification

All primer pairs developed in rainbow trout amplified the expected size amplicons in at least four of the nine other salmonid species. A two-way ANOVA was performed on the frequency of cross-species amplifications as a dependent variable and the unique/duplicated status of the marker and the origin of the microsatellite as factors. The results showed no difference between unique and duplicated loci, which were 87.3 and 86.7% respectively ($F_{1,66} = 0.031$; $P = 0.86$). Microsatellites derived from EST amplified in more species than were anonymous microsatellites at 94% vs. 80%, respectively ($F_{1,66} = 11.4$; $P = 0.001$). No significant interaction effects were detected ($F_{1,66} = 0.015$; $P = 0.47$).

For anonymous microsatellites, 24 of 196 alleles (12%) were common among rainbow trout and at least one of the

Table 1 Source, repeat information, number of alleles, allele size range and polymorphism information content of 35 rainbow trout single-copy microsatellite markers.

cross-species with an average value of 1.4 common allele/locus (0–3). In EST-microsatellites 73 of 301 alleles (24%) were shared by rainbow trout and at least one cross-species with an average of 4.1 common alleles/locus representing (0–7). Principal coordinates analyses were performed on a similarity matrix generated from genotyping data of all single-copy markers of each type. Two three-dimensional PCO plots were drawn showing different relationships between species with respect to markers source (Fig. 2). Anonymous microsatellite data lead to a good separation of fish taxonomic families on a plot accounting for 39.1% of total variance. The *Oncorhynchus* genus is opposed to *Salmo* and *Thymallus* families and to *Salvelinus* according to PCO 2 and PCO 3, respectively. PCO 2 separated *Salvelinus* species from *Salmo* species, the only representative of *Thymallus* being close to *Salmo*. In a plot accounting for 37.4% of total

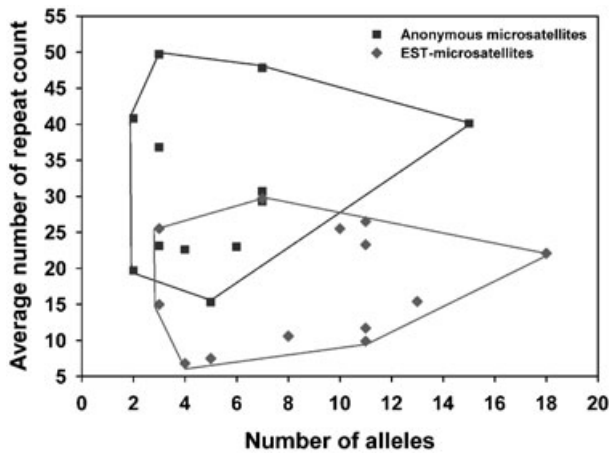


Figure 1 Relationship between the number of alleles and the average dinucleotide repeat counts for anonymous and expressed sequence tag single-copy microsatellites in rainbow trout.

variance, EST-microsatellite data could not clearly separate taxonomic families and members of the same family appeared less close than for anonymous microsatellites.

Table 2 summarizes the frequency of rainbow trout EST and anonymous microsatellites amplifying in each cross-species. On average, 85% of anonymous unique-locus microsatellites amplified in *Oncorhynchus* species while

79% could be transferred to *Salvelinus*, *Salmo* and *Thymallus* altogether. The difference was not significant because of the low value observed in Chinook salmon (65%) which increased the *Oncorhynchus* within-group variance. For transcribed unique-locus microsatellites we observed a significantly higher frequency of amplification within *Oncorhynchus* than in the others families of the cross-species panel, 97 and 89% respectively ($F_{1,7} = 10.8$; $P = 0.01$). No significant differences were observed between *Oncorhynchus* species and the other cross-species for the frequency of transferability of duplicated-locus from genomic source (81% vs. 78%) and from ESTs (97% vs. 96%).

Discussion

The present study was designed to extend available microsatellite resources for rainbow trout. As a by-product of an EST project, microsatellites are isolated for a negligible cost compared with sequencing repeat-enriched genomic libraries. Additionally, microsatellites isolated from EST-databases can be more representative of repeat motifs in a genome than are the enriched genomic microsatellites as enrichment protocols target specific repeats. However, EST-based microsatellite isolation methods ignore repeats located in introns.

Figure 2 Three-dimensional principal coordinates (PCO) analysis plots showing genetic similarity between rainbow trout and nine salmonids cross-species based on 196 alleles derived from 17 anonymous microsatellites (A) and 301 alleles from 18 expressed sequence tag-microsatellites (B). Salmonids families are depicted by different symbols and species are represented as follows: RT, rainbow trout; AC, Arctic char; AS, Atlantic salmon; BnT, Brown trout; BrT, Brook trout; CC, Coastal cutthroat trout; CH, Chinook; CO, Coho; GR, Arctic grayling; SO, Sockeye. Plots A and B accounted for 39.1 and 37.4% of total variance, respectively.

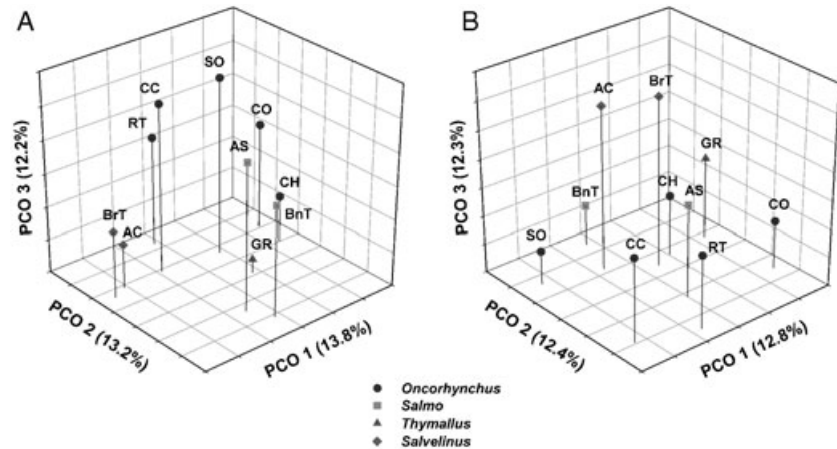


Table 2 Cross-species amplification (%) of rainbow trout EST and genomic microsatellites markers in nine related species.

Locus status	Markers source	<i>Oncorhynchus</i>					Other Salmonids					
		CH	CC	CO	SO	Average	AC	AS	BrT	BnT	GR	Average
Unique locus markers	Genomic (17)	65	94	82	100	85	65	94	88	82	65	79
	ESTs (18)	94	100	100	94	97	83	94	89	89	89	89
Duplicated locus markers	Genomic (20)	70	100	75	80	81	75	90	75	85	65	78
	ESTs (15)	100	100	87	100	97	93	100	93	100	93	96

EST, expressed sequence tag; AC, Arctic char; AS, Atlantic salmon; BnT, Brown trout; BrT, Brook trout; CC, Coastal cutthroat trout; CH, Chinook; CO, Coho; GR, Arctic grayling; SO, Sockeye.

Previous studies have shown that microsatellites in transcribed sequences can be highly polymorphic as observed for *Medicago truncatula* (Eujayl *et al.* 2004), carp (Yue *et al.* 2004) and catfish (Liu *et al.* 1999). The level of polymorphism contained in transcribed microsatellites is determined by the gene region. Microsatellites within UTRs are reported to contain more variability than those in coding regions (Scott *et al.* 2000; Thiel *et al.* 2003). The putative location of our single-copy EST microsatellites in UTRs and their high variability are in agreement with these findings. Our results suggest that the 12 dinucleotide repeat number threshold is efficient in detecting high polymorphic EST-microsatellites. Expressed sequence tag microsatellites variability was associated with a lower number of repeats compared with anonymous microsatellites, as described by Thiel *et al.* (2003) in barley. Several studies have exhibited a positive correlation between genetic diversity and the number of tandem repeats in microsatellite regions (Goldstein & Clark 1995; Jin *et al.* 1996; Innan *et al.* 1997). Nevertheless, unlike non-coding sequence regions, strong evolutionary and functional constraints limit microsatellite repeat expansion in expressed portion of the genome (Dokholyan *et al.* 2000; Metzgar *et al.* 2000). However, intron insertion and variability may interfere in EST microsatellite analysis as was apparently the case for OMM5188.

Cross-species amplification was observed to evaluate marker potential for comparative mapping in other salmonids. Single-copy microsatellite amplification appeared negatively correlated with the phylogenetic distance to rainbow trout. The higher frequency of cross-amplification for EST-microsatellites compared with anonymous microsatellite is in agreement with previous studies and is because of their higher level of conservation among the salmonids. For the same reason, genetic diversity depicted by transcribed microsatellites data only are in disagreement with previous phylogenetic relationships described within salmonid taxa (Oakley & Philips 1999; Crespi & Fulton 2004). Similar observations were made in pine (Liewlaksaneeyanawin *et al.* 2004) and wheat (Gupta *et al.* 2003) where EST-microsatellites failed to differentiate closely related taxa. Still, the phylogenetic relationships among the salmonids determined by our data should be confirmed on larger cross-species sample sizes.

According to Bailey *et al.* (1978), several groups of fishes including salmonids and catostomids have retained approximately 50% of their loci as duplicates over the 50 million years since tetraploidy occurred. This is consistent with our findings. Additionally, our data also confirmed a certain degree of conservation for this part of the genome within the salmonid fishes. Duplicated microsatellite transferability to other salmonids was not correlated with the phylogenetic distance suggesting that these markers evolved at slower rate than the single-copy microsatellites. Divergence of both duplicates of the same locus can make these markers relevant for mapping and

for investigating and inferring the evolutionary history of duplication among the salmonids (Nichols *et al.* 2003; Woram *et al.* 2004). Their use for population genetics is conditioned by a high degree of differentiation between duplicates that does not allow common alleles between duplicates.

In conclusion, ESTs in public databases are valuable resources for developing Type I microsatellite markers which can be used for population and evolutionary genetic studies and have great potential for developing comparative maps. Additionally, the costs associated with EST-microsatellite development are negligible compared with repeat-enriched library protocol. Further characterization of the genes they are associated with will offer opportunity to investigate the effect of repeat polymorphism on gene function. Their high level of interspecific transferability compensates their low ability to address phylogenetic relationship in salmonids compared with microsatellites derived from repeat-enriched libraries.

Acknowledgements

We thank Roseanna Athey, Kristy Anderson and Renee Fincham for technical assistance. We are grateful to Dr Tim King (USGS, Kearneysville, WV, USA), Dr Jim Seeb (Alaska Department of Fish and Game, Anchorage, AK), Dr Gary Thorgaard (Washington State University, Pullman, WA), Dr Linda Park (National Oceanic and Atmospheric Administration, Seattle, WA), James Gray (Wolf Creek National Fish Hatchery, KY) for the supply of tissue/DNA materials, and Dr Cheryl Morrison (USGS, Kearneysville, WV, USA) for critical review of this manuscript.

References

- Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–10.
- Anderson J.A., Churchill G.A., Autrique J.E., Tanksley S.D. & Sorrells M.E. (1993) Optimizing parental selection for genetic linkage maps. *Genome* **36**, 181–86.
- Arden W.R. & Kapuscinski A.R. (2003) Demographic and genetic estimates of effective population size (N_e) reveals genetic compensation in steelhead trout. *Molecular Ecology* **12**, 35–49.
- Bailey G.S., Poulter R.T.M. & Stockwell P.A. (1978) Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. *Proceedings of the National Academy of Science of the United States of America* **75**, 5575–9.
- Benson G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* **27**, 573–80.
- Botstein D., White R.L., Skolnick M. & Davis R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314–31.
- Boutin-Ganache I., Raposo M., Raymond M. & Deschepper C.F. (2001) M13-tailed primers improve the readability and usability

- of microsatellite analyses performed with two different allele-sizing methods. *Biotechniques* **31**, 24–6, 28.
- Cho Y.G., Shii T., Temnyk S., Chen X., Lipovich L., McCouch S.R., Park W.D., Ayres N. & Cartinhour S. (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **100**, 713–22.
- Colson I. & Goldstein D.B. (1999) Evidence for complex mutation at microsatellite in *Drosophila*. *Genetics* **152**, 617–27.
- Crespi B.J. & Fulton M.J. (2004) Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. *Molecular Phylogenetics and Evolution* **31**, 658–79.
- Dokholyan N.V., Buldyrev S.V., Havlin S. & Stanley H.E. (2000) Distributions of dimeric tandem repeats in non-coding and coding DNA sequences. *Journal of Theoretical Biology* **202**, 273–82.
- Eujayl I., Sledge M.K., Wang L., May G.D., Chekhovskiy K., Zwonitzer J.C. & Mian M.A. (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theoretical and Applied Genetics* **108**, 414–22.
- Goldstein D.B. & Clark A.G. (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Research* **23**, 3882–6.
- Gupta P.K., Rutsge S., Sharma S., Singh R., Kumar N. & Balyan H.S. (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics* **270**, 315–23.
- Haddad L.A., Fuzikawa A.K. & Pena S.D. (1997) Simultaneous detection of size and sequence polymorphisms in the transcribed trinucleotide repeat D2S196E (EST00493). *Human Genetics* **99**, 796–80.
- Hammer R., Harper D.A.T. & Ryan P.D. (2001) PAST: Paleontological Statistics Software Package for education and data analysis. *Palaeontologia Electronica* **4**, 1–9.
- Heath D.D., Busch C., Kelly J. & Atagi D.Y. (2002) Temporal change in genetic structure and effective population size in steelhead trout (*Oncorhynchus mykiss*). *Molecular Ecology* **11**, 197–214.
- Innan H., Terauchi R. & Miyashita T. (1997) Microsatellites polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. *Genetics* **146**, 1441–52.
- Jin L., Macaubas C., Hallmayer J., Kimura A. & Mignot E. (1996) Mutation rate varies among alleles at a microsatellites locus: phylogenetic evidence. *Proceedings of the National Academy of Science of the United States of America* **93**, 15285–8.
- Liewlaksaneeyanawin C., Ritland C.E., El-Kassaby Y.A. & Ritland K. (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theoretical and Applied Genetics* **109**, 361–9.
- Liu Z., Tan G., Li P. & Dunham R.A. (1999) Transcribed dinucleotides microsatellites and their associated genes from Channel catfish *Ictalurus punctatus*. *Biochemical and Biophysical Research Communications* **259**, 190–4.
- MacCrimmon H.R. (1971) World distribution of rainbow trout (*Salmo gairdneri*). *Journal of the Fisheries Research Board of Canada* **28**, 663–704.
- Metzgar D., Bytof J. & Wills C. (2000) Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Research* **10**, 72–80.
- Nichols K.M., Young W.P., Danzmann R.G. *et al.* (2003) A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics* **34**, 102–15.
- Nielsen J.L., Crow K.D. & Fountain M.C. (1999) Microsatellite diversity and conservation of a relic trout population: McCloud River redband trout. *Molecular Ecology* **8**, S129–42.
- O'Brien S.J., Womack J.E., Lyons L.A. *et al.* (1993) Anchored reference loci for comparative genome mapping in mammals. *Nature Genetics* **3**, 103–12.
- Oakley T.H. & Phillips R.B. (1999) Phylogenies of salmonine fishes based on growth hormone introns: Atlantic salmon (*Salmo*) and Pacific (*Oncorhynchus*) are not sister taxa. *Molecular Phylogenetics and Evolution* **11**, 381–93.
- Ortí G., Pearce D.E. & Avise J.C. (1997) Phylogenetic assessment of length variation at a microsatellite locus. *Proceedings of the National Academy of Science of the United States of America* **94**, 10745–9.
- Rychlik W. & Rhoads R.E. (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Research* **17**, 8543–51.
- Sakamoto T., Danzmann R.G., Gharbi K. *et al.* (2000) A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* **155**, 1331–45.
- Sambrook J. & Russell D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Scott K.D., Eggler P., Seaton G., Rossetto M., Ablett E.M., Lee L.S. & Henry R.J. (2000) Analysis of SSRs derived from grape ESTs. *Theoretical and Applied Genetics* **100**, 723–6.
- Serapion J., Kucuktas H., Feng J. & Liu Z. (2004) Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Marine Biotechnology* **6**, 364–77.
- Silverstein J.T., Rexroad C.R. & King T.L. (2004) Genetic variation measured by microsatellites among three strains of domesticated rainbow trout (*Oncorhynchus mykiss*, Walbaum). *Aquaculture Research* **35**, 40–8.
- Thiel T., Michalek W., Varshney R.K. & Graner A. (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* **106**, 414–22.
- Venkatesh B. (2003) Evolution and diversity of fish genomes. *Current Opinion in Genetics and Development* **13**, 588–92.
- Woram R.A., McGowan C., Stout J.A. *et al.* (2004) A genetic linkage map for Arctic char (*Salvelinus alpinus*): evidence for higher recombination rates and segregation distortion in hybrid versus pure strain mapping parents. *Genome* **47**, 304–15.
- Young W.P., Wheeler P.A., Fields R.D. & Thorgaard G.H. (1996) DNA fingerprinting confirms isogenicity of androgenetically derived rainbow trout lines. *Journal of Heredity* **87**, 77–80.
- Yue G.H., Ho M.Y., Orban L. & Komen J. (2004) Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. *Aquaculture* **234**, 85–98.